

# Simple Mathematical Model Of Pathologic Microsatellite Expansions: When Self-Reparation Does Not Work

Boris Veytsman

*George Mason University, School of Computational Sciences, MS 5C3, Fairfax,  
VA 22030, USA, E-mail: borisv@lk.net*

Leila Akhmadeyeva

*Bashkir State Medical University, 3 Lenina Str., Ufa, 450077, Russia. E-mail:  
medic@agidel.ru*

---

## Abstract

We propose a simple model of pathologic microsatellite expansion, and describe an inherent self-repairing mechanism working against expansion. We prove that if the probabilities of elementary expansions and contractions are equal, microsatellite expansions are always self-repairing. If these probabilities are different, self-reparation does not work. Mosaicism, anticipation and reverse mutation cases are discussed in the framework of the model. We explain these phenomena and provide some theoretical evidence for their properties, for example the rarity of reverse mutations.

*Key words:* Microsatellite Expansion, Myotonic Dystrophy, Huntington Disease, Fragile X, Mathematical Model

---

## 1 Introduction

Pathologic microsatellite expansion is a phenomenon causing several severe diseases like Fragile X, Huntington disease, Myotonic Dystrophy and others (Harper, 2001; Pearson, 2003; Piñeiro et al., 2003; Libby et al., 2003; Pearson et al., 2005). There are places in a DNA molecule where nucleotide sequences are repeated several times. The number of such repeats (satellites) is usually stable during normal replication. However, sometimes a mutation occurs, and the mutated DNA has more (expansion) or less (contraction) repeats than its ancestor. Normally the mutation rates are about  $10^{-3} \dots 10^{-4}$  per generation per locus (Ellegren, 2000b). However in the case of diseases mentioned above, the

expansions occur much faster, at a rate of hundreds or thousands per locus per generation. We will call this phenomenon *pathologic expansion* to distinguish it from the much slower “normal” expansion. We are interested in the case where the number of nucleotides in the repeating sequence is small, e.g. when the repeated sequences are triplets. In this case the phenomenon is usually called microsatellite expansion. Sometimes, as in the case of Myotonic Dystrophy Type 1 (DM1, OMIM #160900), the number of repeated triplets might actually reach thousands. For some diseases this expansion occurs in a coding part of DNA, for some in a non-coding one, but it is always a multi-system disease with multiple symptoms.

There are several notable features of pathologic microsatellite expansion, common to most diseases associated with it:

**Mosaicism:** For most diseases the number of repeats is *not* the same in all cells. Rather, it has a wide distribution of possible values. A notable exception is Huntington disease (HD, OMIM #143100), where the mosaicism is not as prominent as for other expansion-related diseases (Harper and Jones, 2002). The reason for becomes more clear after we discuss the model. We return to this disease in Section 4.

**Anticipation:** For some diseases a relatively small increase in the number of repeats does not lead to symptoms. However, the stability of the repeated sequence is lower than the stability of the non-affected DNA, and the children of affected parents might show symptoms, sometimes severe (Harper, 2001).

**Reverse Mutation:** Sometimes the children of symptomatic patients have the normal number of repeats. This is a rare, but still observable phenomenon (Brunner et al., 1993; Monckton et al., 1995).

A theory of microsatellite expansion must naturally explain these phenomena.

One of the most common explanations of microsatellite expansion is the formation of hairpins either during replication (Cleary et al., 2002; Mirkin and Smirnova, 2002; Pearson, 2003; Yang et al., 2003) or during DNA repair after transcription (Gomes-Pereira et al., 2004). In both these cases the hairpin formation can cause either expansion or contraction of DNA. A recent review comparing these explanations can be found in, e.g. (Pearson et al., 2005). In this paper we will not try to distinguish between these mechanisms; the proposed model describes both. Therefore we will understand by *cell events* either cell divisions in the first model or cell repair events in the second one.

This model is attractive because it can explain a number of features of microsatellite expansion. In this model hairpins form during some, but not all, cell events. Therefore it is a random, rather than a deterministic process. Thus different cells have different number of expansions and contractions in their

histories. This explains mosaicism, i.e. broad distribution of the number of repeats in different cells within the same tissue.

Gametes in this model might have different numbers of repeats. If the number of repeats turns out to be small, the child of an affected parent will be not affected. This explains reverse mutation. The calculations below show that this phenomenon is indeed very rare.

To explain anticipation, we can assume that the probability of expansion grows with the number of repeats in the DNA molecule. If the number of repeats in an asymptomatic patient increases, the probability to have a symptomatic child also increases.

One of the ways to verify these speculations is to try to make possible conclusions from the model, and to check whether these conclusions agree with the observed picture of microsatellite expansion. If they do, our confidence in the model grows, if they do not, then the model is wrong. This paper takes the qualitative model described above for granted and tries to formalize it in the form of differential equations for the observed distribution of repeats. We solve these equations and show a qualitative agreement with the observations.

It is interesting to compare the pathologic microsatellite expansion due to fast mutations with the “normal” microsatellite expansion due to slow mutations. The latter got much attention (see e.g. (Di Rienzo et al., 1994; FitzSimmons et al., 1995; Goldstein et al., 1995; Angers and Bernatchez, 1997; Primmer and Ellegren, 1998; Schlötterer et al., 1998; Brinkmann et al., 1998; Makova et al., 2000; Kayser et al., 2000; Ellegren, 2000a; Huang et al., 2002; Calabrese and Durrett, 2003) and the review (Ellegren, 2000b)) as a way to infer data on evolution process. This approach has a promise of higher time resolution than other methods because “slow” microsatellite mutations are still several orders of magnitude faster than most other mutations (Ellegren, 2000b). Such full comparison of “normal” and pathologic expansions is beyond this paper, but one might express a hope that in the future it will help to understand both better. For example, while during “normal” mutations the expansions of the repeat sequence are thought to be more probable than the contractions, there seems to exist some mechanism that limits an uncontrolled expansion of microsatellites (Garza et al., 1995; Amos and Rubinsztein, 1996; Harr and Schlötterer, 2000; Xu et al., 2000). Apparently such mechanism is absent or too weak for pathologic microsatellite expansions. A comparative study might therefore help to elucidate details and effects of this mechanism. On the other hand, one must be very cautious in the application of data and conclusions from the study of the “normal” expansions to the pathologic ones and vice versa.

There are many theoretical works describing “normal” microsatellite expansion using both analytical methods and computer simulations (see e.g. (Tachida and Iizuka,

1992; Shriver et al., 1993; Nielsen, 1997; Bell and Jurka, 1997; Kruglyak et al., 1998, 2000; Calabrese et al., 2001; Sibly et al., 2001; Whittaker et al., 2003; Shinde et al., 2003; Sibly et al., 2003; Calabrese and Durrett, 2003; Lai et al., 2003; Lai and Sun, 2003a,b)). However, they deal with slowly changing sequences with relatively small number of repeats. Our situation is rather opposite: we are interested in long sequences and fast change. Therefore our formalism and results are quite different from theirs.

## 2 Model

First, let us discuss how one hairpin is formed. Consider a hairpin with  $h/2$  repeats. If  $l_k$  is the number of repeats in a Kuhn segment of the polymer (de Gennes, 1979; Painter and Coleman, 1997), then we gain  $h/(2l_k)$  degrees of freedom. The corresponding free energy loss is  $kTh/l_k$ , where  $k$  is Boltzmann constant,  $T$  is temperature. On the other hand the energy gain is  $\delta E h$ , where  $\delta E$  is the energy gain per repeat. Since both these contributions are proportional to  $h$ , the total free energy change is also proportional to  $h$ :

$$\Delta F = Ch, \quad C = \text{const} \quad (1)$$

The value of the constant  $C$  in this equation depends on the relation between  $\delta E$  and  $kT/l_k$ . If  $C < 0$ , the formation of hairpins causes a decrease of free energy, and the longer are the hairpins, the better. This would lead to a fast de-stabilization of the number of repeats. Since this does not happen, we can conclude that  $C > 0$ . This means that the formation of hairpins is *not* encouraged by thermodynamics, and the formation of longer hairpins is suppressed with the probability proportional to  $\exp(-Ch)$ . Since the probability exponentially decreases with  $h$ , only the shortest possible hairpins are formed. The minimal size of a hairpin depends on the flexibility of the molecule. It stands to reason to assume it of the order of one-two Kuhn segments. Therefore the microsatellite de-stabilization cannot start until the DNA has at least several  $l_k$  repeats.

These thermodynamic considerations explain anticipation: it is necessary to have at least several Kuhn segments in the microsatellite repeats interval to start the mechanism of de-stabilization. Of course *cis*-elements might subtly influence hairpin formation at the early stages of de-stabilization. Therefore they play an important role in the transition from anticipation to disease (Brock et al., 1999; Cleary et al., 2002). It is interesting that a certain threshold number of repeats is necessary for “normal” expansions too, at least in some cases (Sibly et al., 2001, 2003; Shinde et al., 2003).

Let us now discuss a strand of DNA having  $x$  repeats after  $i$  cell events. The

next event can have one of three possible outcomes:

- (1) No expansion or contraction occurred.
- (2) There was an expansion of length  $n$ . Let  $Q_{\text{ins}}(x, n)$  be the probability of this event.
- (3) There was a contraction of length  $n$ . Let  $Q_{\text{del}}(x, n)$  be the probability of this event.

Let  $P_i(x)$  be the probability that the strand has exactly  $x$  repeats. Then it is easy to write the master equation, describing the transition from the step number  $i$  to the step number  $i + 1$ :

$$P_{i+1}(x) = P_i(x) + \sum_{n=1}^{\infty} \left( P_i(x-n)Q_{\text{ins}}(x-n, n) + P_i(x+n)Q_{\text{del}}(x+n, n) - P_i(x)Q_{\text{ins}}(x, n) - P_i(x)Q_{\text{del}}(x, n) \right) \quad (2)$$

Equation (2) might be simplified if we make the following assumption, based on the thermodynamic considerations in the beginning of this Section. Namely, we assume that the constant  $C$  in equation (1) is large enough, so expansions and contractions are in fact rare. If  $n_{\text{min}}$  is the minimal hairpin length allowed by chain flexibility, then the only events to be considered in the sum (2) are expansions and contractions of length  $n_{\text{min}}$ . Now we must estimate the probabilities of one expansion or contraction as functions of repeats number  $x$ . If we consider for guidance “slow” mutations in the non pathologic regime (see Introduction), we see that there is a considerable controversy in the literature about the dependence of mutation rate on  $x$ . Some authors report exponential growth (Brinkmann et al., 1998; Whittaker et al., 2003; Lai and Sun, 2003a), while other report much weaker linear relationship (Kruglyak et al., 1998, 2000; Sibly et al., 2001; Shinde et al., 2003) or more complex dependence (Calabrese and Durrett, 2003; Sibly et al., 2003). Moreover, *cis*-factors, obviously, should also influence the mutation rates. We can only agree with Primmer et al. (1998): “These observations demonstrate that the mutation process of microsatellites may be more complex than previously thought”. Fortunately the situation for relatively large  $x$  can be simplified. Indeed, if the number of repeats is sufficiently large, we can divide the stretch of microsatellites into parts, each enough to assume that a mutation or a repair error in one part does not affect the other ones. Only two end parts depend on the *cis*-factors. If each part mutates independently, the overall mutation rate should be proportional to the number of parts. This simple consideration suggests that the mutation rates at least for large  $x$  should be linear in  $x$ . Moreover, they must go to zero as the number of repeats goes to  $n_{\text{min}}$ . If we

set the origin of  $x$  axis to  $n_{\min}$ , we get simply:

$$Q_{\text{ins}}(x, n) = q_{\text{ins}}(n)x, \quad Q_{\text{del}}(x, n) = q_{\text{del}}(n)x \quad (3)$$

With these assumptions equation (2) can be rewritten as:

$$\begin{aligned} P_{i+1}(x) - P_i(x) = & \\ & q_{\text{ins}}(n_{\min}) \left( P_i(x - n_{\min})(x - n_{\min}) - P_i(x)x \right) + \\ & q_{\text{del}}(n_{\min}) \left( P_i(x + n_{\min})(x + n_{\min}) - P_i(x)x \right) \end{aligned} \quad (4)$$

The next step is the transition from the discrete representation (4) to a continuous one. We will “smooth” the variables  $i$  and  $x$ . In order to do this we will measure “time”  $t$  in the number of events and consider  $P$  to be a function of a continuous variables  $t$  and  $x$ , so  $P(t, x) dx$  is the probability to have the number of repeats between  $x$  and  $x + dx$  at the time  $t$ . Then we can rewrite equation (4) in the continuous form as:

$$\begin{aligned} \frac{\partial P(t, x)}{\partial t} &= -c \frac{\partial (xP(t, x))}{\partial x} + D \frac{\partial^2 (xP(t, x))}{\partial x^2} \\ c &= (q_{\text{ins}}(n_{\min}) - q_{\text{del}}(n_{\min})) n_{\min} \\ D &= \frac{q_{\text{ins}}(n_{\min}) + q_{\text{del}}(n_{\min})}{2} n_{\min}^2 \end{aligned} \quad (5)$$

Note that the quantity

$$J = cxP - D \frac{\partial (xP)}{\partial x} \quad (6)$$

has the meaning of *flow* of probability through the point  $x$  at the time  $t$ . By the way, this means that equation (5) has the simple meaning of continuity equation  $\partial P / \partial t + \text{div } J = 0$ .

If the number of repeats in the zygote is  $x_0$ , then equation (5) has the following initial condition:

$$P(0, x) = \delta(x - x_0) \quad (7)$$

where  $\delta$  is Dirac’s delta-function (e.g. Barton, 1989).

We will see that  $P(t, +0)$  remains finite, so

$$\lim_{x \rightarrow 0} xP(t, x) = 0 \quad (8)$$

As shown below, the flow (6) remains non-zero at  $x \rightarrow +0$ . Therefore the integral

$$f_m(t) = \int_{+0}^{\infty} P(t, x) dx \quad (9)$$

is *not* conserved. This integral represents the fraction of “mutant” cells, i.e. cells with the number of repeats large enough to form hairpins and therefore to be described by equation (5). The discontinuity of the function  $P(t, x)$  at  $x \rightarrow 0$  makes this integral less than 1. Its complement to 1 is the fraction of cells, which can no longer form hairpins and are “stuck at zero”:

$$f_r(t) = 1 - f_m(t) \quad (10)$$

We will call such cells “repaired” cells. The increase of  $f_r$  over time represents a self-reparation effect.

We introduce the parameter

$$\gamma = \frac{cx_0}{D} \quad (11)$$

This parameter reflects the difference between the probabilities of expansion and contraction. The case of  $\gamma = 0$  corresponds to the situation when expansions and contractions occur with equal probabilities. If expansions are more probable, then  $\gamma > 0$ . Note that the value of  $\gamma$  depends on the progenitor number of repeats  $x_0$ . The greater is  $x_0$ , the larger is  $\gamma$ . We will see that this parameter critically affects the microsatellite instability.

We will measure time in the units of  $x_0/D$ , i.e. we will introduce a dimensionless variable

$$\tau = tD/x_0 \quad (12)$$

As shown in Appendix, at the reasonable values for the parameters one dimensionless unit of time corresponds to about 25 cell events.

### 3 Results And Discussion

The solution for equation (5) is obtained in Appendix A. Here we discuss the properties of the solution and predictions of the model.

First we consider the fraction of repaired cells (see Appendix A):

$$f_r(\tau) = \exp\left(-\frac{\gamma}{1 - e^{-\gamma\tau}}\right) \quad (13)$$

The number of “repaired” cells increases with the time  $\tau$ . The speed of this increase and the limit fraction at  $\tau \rightarrow \infty$  depend on the parameter  $\gamma$ .

In the special case  $\gamma = 0$ , i.e. when expansions and contractions happen with the same probability, equation (13) becomes  $f_r(\tau) = \exp(-1/\tau)$ . In this case  $f_r$  goes to 1 as  $\tau \rightarrow \infty$ . This means that all cells eventually become repaired.

In the case  $\gamma > 0$  the limit  $f_r$  at  $\tau \rightarrow \infty$  is  $\exp(-\gamma)$ . For large enough  $\gamma$  the fraction of repaired cells is small, but nevertheless not zero. This case corresponds to the observed clinical picture.

Now we can explain the phenomenon of reverse mutation. If a parent is affected, the gamete might carry DNA either from the repaired population or from unrepaired, mutant population. In the first case a reverse mutation occurs. Therefore the probability of reverse mutation is  $\exp(-\gamma)$ . It seems that reverse mutations are very rare events. In the case of Myotonic Dystrophy (Brunner et al., 1993) the probability of reverse mutation is very small. We will rather arbitrarily estimate it as 1 : 1000; a more frequent occurrence would be observed more often, and a more rare one would not be observed at all. This gives the following estimate for  $\gamma$ :

$$\gamma \approx 7 \quad (14)$$

Plots of  $f_r(\tau)$  for several values of  $\gamma$  are shown on Figure 1. It can be seen from these plots that the number of repaired cells quickly reaches the limit value. This justifies the assumption that the fraction of repaired cells in the gametes is equal to the limiting value.

Let us now return to the solution of equation (5). It can be expressed through the mean number of repeats  $m$  and standard deviation  $\sigma$  (see Appendix A):

$$P(x, t) = \frac{2m^{3/2}}{\sigma^2 x^{1/2}} \exp \left[ -\frac{2m^2}{\sigma^2} \left( 1 + \frac{x}{m} \right) \right] I_1 \left( \frac{4m^{3/2} x^{1/2}}{\sigma^2} \right) \quad (15)$$

where  $I_1$  is the modified Bessel function (Abramowitz and Stegun, 1972, § 9). The mean number of repeats and standard deviation depend on time as

$$m = x_0 \exp(\gamma\tau), \quad \sigma = \left( 2 \frac{1 - \exp(-\gamma\tau)}{\gamma} \right)^{1/2} x_0 \exp(\gamma\tau) \quad (16)$$

Also interesting are skewness and kurtosis of the distribution. They are

$$S = 3 \left( \frac{1 - \exp(-\gamma\tau)}{2\gamma} \right)^{1/2}, \quad K = 6 \frac{1 - \exp(-\gamma\tau)}{\gamma} \quad (17)$$

At early stage of instability growth ( $\gamma\tau \ll 1$ ) these equations describe a sharp distribution centered around  $m$ . The ratio of the distribution width  $2\sigma$  to the mean size  $m$  is small (about  $2(2\tau)^{1/2}$ ).

However, at later stages ( $\gamma\tau \gg 1$ ) the picture is completely different. At these stages the curve is very wide. The ratio of the width to the the mean size is



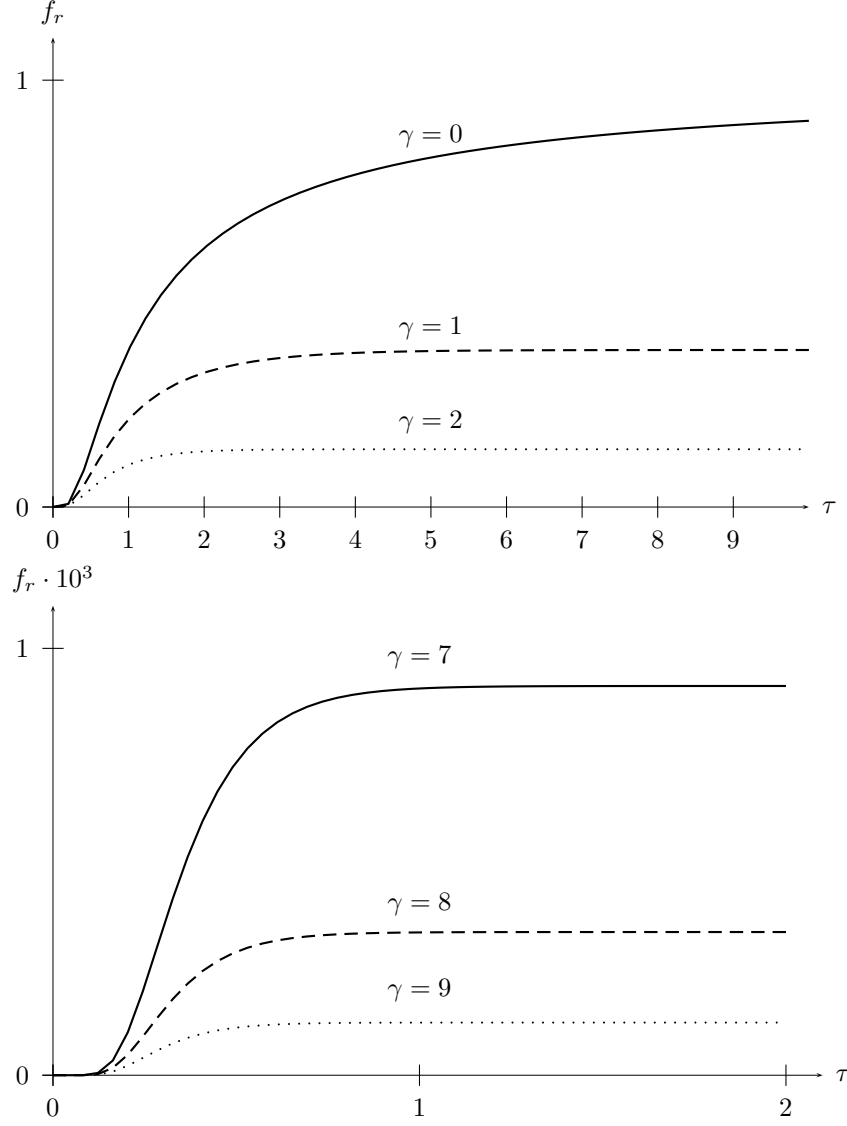


Fig. 1. Fraction of Repaired Cells As Function of Time

at these stages  $2\sigma/m = 2(2/\gamma)^{1/2} \approx 1$ . This large width explains the observed mosaicism.

The transition from the first regime to the second one depends on  $\gamma$ , and thus on the progenitor number of repeats  $x_0$ . The larger is  $x_0$ , the earlier is the transition to the second regime, i.e. the regime of developed instability.

The distribution has positive skewness and kurtosis. They are about zero at early stages, and tend to  $3/(2\gamma)^{1/2} \approx 0.8$  and  $6/\gamma \approx 0.9$  correspondingly.

Typical plots distribution of repeat lengths are shown on Figures 2 and 3.

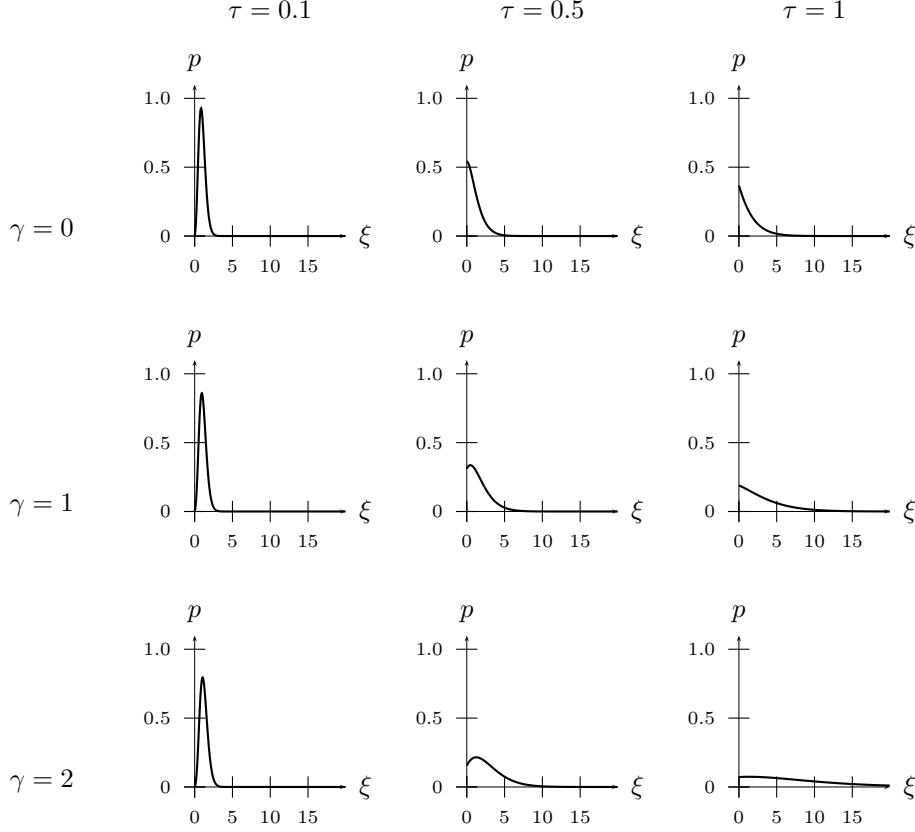


Fig. 2. Repeat Number Distribution For Unrepaired Cells, small  $\gamma$

#### 4 Conclusions

We have shown that a very simple model of pathologic microsatellite expansion can qualitatively explain the observed phenomena of anticipation, mosaicism and spontaneous recovery. This model considers expansion or contraction of repeats as a random process with the probability of expansion and contraction related to the probability of hairpin formation. A mathematical model based on this picture is able to predict the shape of the distribution of the number of repeats after many divisions.

This model predicts a natural “reparation process” leading to reverse mutation. In the case when the probabilities of expansion and contraction are equal, this process eventually heals the mutation. Therefore mutation survives only if the probability of expansion exceeds the probability of contraction. The fraction of repaired cells in the long run depends on this difference in probabilities.

We implicitly assumed that only “young” cells, the ones belonging to the latest generation, are used in the measurements of the number of repeats. If this assumption is not satisfied, the observed length distribution should be obtained by summation of the results over generations of cells. However, if

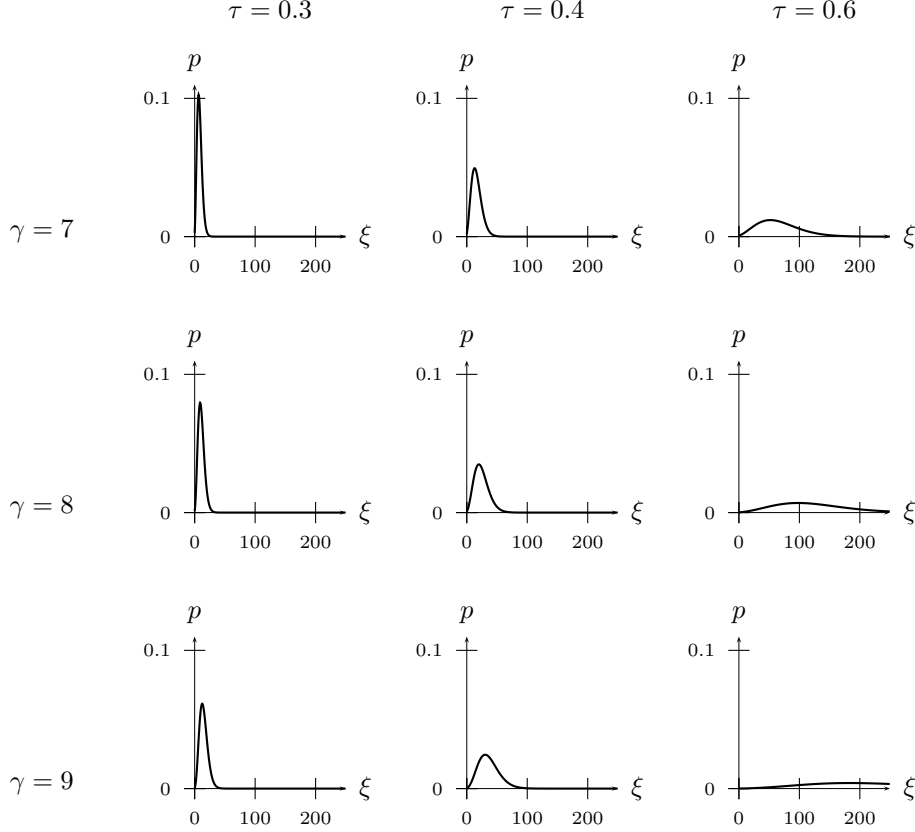


Fig. 3. Repeat Number Distribution For Unrepaired Cells, large  $\gamma$

the “older” cells die due to apoptosis, this effect is small. For example, if the apoptosis for blood cells occurs after 25–40 mitoses, as it is usually thought, then the effect is indeed negligible for blood samples.

Another interesting question is the possibility of selection: the rate of cell survival and multiplication might depend on the level of the mutation of microsatellite expansion. This will change the rates of expansion and contraction for the cell population as a whole.

A notable exception from the general picture of trinucleotide expansion diseases is Huntington disease. Mosaicism for this disease is not as prominent as for other dynamic mutations (Harper and Jones, 2002). However, a closer look shows that this example actually does not contradict our model. The number of repeats for HD is rather small (several dozens). It seems that the mutation in this case is caused by a small number of relatively large expansions, rather than a large number of small expansions, as assumed in this paper.

It would be interesting to extend the analysis of this paper to quantitative comparison with the experimental data. This will be done in subsequent works.

A further comparison of the fast pathologic mutations and slow “normal” ones

seems also to be promising.

## Acknowledgements

The authors are grateful to R. Colby, C.E. Pearson, R. Korneluk, K. Makova, D. Monckton, Y. Giguere and F. Rousseau for helpful discussions. This work was supported by International Research and Exchange Board (IREX, 2003 ECA Alumni Small Grant), International Association for the Promotion of Cooperation with Scientists from the New Independent States of the Former Soviet Union (INTAS, grant YSF 00-34), Russia President Grant Program for Young Scientists (grant MD-346.2003.04), Open Society Institute (Soros Foundation) and American Austrian Foundation (AAF).

## A Solution Of Master Equation

In this Appendix we provide the solution of master equation (5). This equation is easier to solve in the following dimensionless variables:

$$\xi = x/x_0, \quad \tau = tD/x_0, \quad p(\tau, \xi) = x_0 P(x, t) \quad (\text{A.1})$$

Let us roughly estimate these parameters. Taking values approximating the known data about DM1, we get

$$x_0 \approx 10^2, \quad q_{\text{ins}} \approx q_{\text{del}} \approx 10^{-2}, \quad n_{\text{min}} \approx 20 \quad (\text{A.2})$$

so

$$\xi \approx 10^{-2}x, \quad \tau \approx 0.04t \quad (\text{A.3})$$

In other words, one dimensionless unit of  $\tau$  corresponds approximately to 25 cell events, while one dimensionless unit of  $\xi$  corresponds approximately to 100 repeats.

Let us introduce the function

$$v(\tau, \xi) = \xi p(\tau, \xi) \quad (\text{A.4})$$

Then equations (5) can be rewritten as

$$\frac{\partial v}{\partial \tau} + \gamma \xi \frac{\partial v}{\partial \xi} - \xi \frac{\partial^2 v}{\partial \xi^2} = 0 \quad (\text{A.5})$$

We use Laplace transform with respect to  $\xi$ :

$$\mathcal{V}(\tau, s) = \int_0^\infty e^{-s\xi} v(\tau, \xi) d\xi \quad (\text{A.6})$$

For the Laplace transform of the derivatives we have (see (Abramowitz and Stegun, 1972, § 29.2.5))

$$\frac{\partial v}{\partial \xi} \doteq s\mathcal{V} - v(t, +0), \quad \frac{\partial^2 v}{\partial \xi^2} \doteq s^2\mathcal{V} - sv(t, +0) - V(\tau) \quad (\text{A.7})$$

where  $\doteq$  means Laplace transform, and

$$V(\tau) = \left. \frac{\partial v(\tau, \xi)}{\partial \xi} \right|_{\xi \rightarrow +0} \quad (\text{A.8})$$

Multiplication by  $-\xi$  corresponds to differentiation by  $s$  (see (Abramowitz and Stegun, 1972, § 29.2.9)). Due to the border condition (8), the values of  $v(t, +0)$  and  $V(\tau)$  in equation (A.7) do not depend on  $\xi$ . Therefore we can rewrite equation (A.5) in Laplace space as

$$\frac{\partial \mathcal{V}}{\partial \tau} + \frac{\partial}{\partial s} (s^2 \mathcal{V} - \gamma s \mathcal{V}) = 0 \quad (\text{A.9})$$

with the initial condition from equation (7)

$$\mathcal{V}(0, s) = \exp(-s) \quad (\text{A.10})$$

Let

$$\mathcal{U}(\tau, s) = (s^2 - \gamma s) \mathcal{V}(\tau, s) \quad (\text{A.11})$$

Then we can rewrite equations (A.9) and (A.10) as

$$\frac{\partial \mathcal{U}}{\partial \tau} + (s^2 - \gamma s) \frac{\partial \mathcal{U}}{\partial s} = 0 \quad (\text{A.12})$$

$$\mathcal{U}(0, s) = (s^2 - \gamma s) \exp(-s) \quad (\text{A.13})$$

The solution of differential equation (A.11) is

$$\mathcal{U}(\tau, s) = f \left( \frac{e^{-\gamma \tau} (s - \gamma)}{s} \right) \quad (\text{A.14})$$

Taking into account the initial condition (A.13) and returning to the function  $\mathcal{V}$ , we get

$$\mathcal{V}(\tau, s) = \frac{e^{-\gamma \tau} \gamma^2}{(s(1 - e^{-\gamma \tau}) + e^{-\gamma \tau} \gamma)^2} \exp \left( -\frac{s\gamma}{s(1 - e^{-\gamma \tau}) + e^{-\gamma \tau} \gamma} \right) \quad (\text{A.15})$$

The reverse Laplace transform of this expression is (see (Abramowitz and Stegun, 1972, § 29.3.81))

$$v = \frac{\gamma (e^{-\gamma \tau} \xi)^{1/2}}{1 - e^{-\gamma \tau}} \exp \left( -\frac{\gamma(1 + e^{-\gamma \tau} \xi)}{1 - e^{-\gamma \tau}} \right) I_1 \left[ \frac{2\gamma}{1 - e^{-\gamma \tau}} (e^{-\gamma \tau} \xi)^{1/2} \right] \quad (\text{A.16})$$

which provides the solution of the master equation.

Using the asymptotic (Abramowitz and Stegun, 1972, § 9.6.7)

$$I_1(z) \approx z/2, \quad z \ll 1 \quad (\text{A.17})$$

we see that at  $\xi \rightarrow 0$

$$p(\tau, +0) = \frac{\gamma^2 e^{-\gamma\tau}}{(1 - e^{-\gamma\tau})^2} \exp\left(-\frac{\gamma}{1 - e^{-\gamma\tau}}\right) \quad (\text{A.18})$$

The flow at  $\xi \rightarrow +0$  is non-zero. The total fraction of repaired cells can be calculated by calculating the integral of flow. Rewriting equation (6) in dimensionless coordinates, we see that

$$f_r(\tau) = \int_0^\tau p(u, 0) du = \exp\left(-\frac{\gamma}{1 - e^{-\gamma\tau}}\right) \quad (\text{A.19})$$

which gives equation (13).

The momenta of function  $p(\xi)$  are defined as

$$\mu'_n(\tau) = \int_0^\infty \xi^n p(\tau, \xi) d\xi, \quad n = 0, 1, 2, \dots \quad (\text{A.20})$$

Laplace transform gives

$$\mu'_n(\tau) = (-1)^{n-1} \lim_{s \rightarrow +0} \frac{\partial^{n-1} \mathcal{V}(\tau, s)}{\partial s^{n-1}}, \quad n = 1, 2, \dots \quad (\text{A.21})$$

Differentiating the function  $\mathcal{V}$ , calculating central momenta (see (Abramowitz and Stegun, 1972, § 26)) and returning to dimensional coordinates, we obtain equations (16) and (17) for mean, deviation, skewness and kurtosis. After transformation of equation (A.16) to dimensional coordinates and substitution of equations (16), we obtain equation (15).

## References

- M. Abramowitz and I. A. Stegun, editors. *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*. Dover Publications, Inc., New York, 1972.
- William Amos and David C. Rubinsztein. Microsatellites are subject to directional evolution. *Nat. Genet.*, 12:13–14, 1996.
- Bernard Angers and Louis Bernatchez. Complex evolution of a salmonid microsatellite locus and its consequences in inferring allelic divergence from size information. *Mol. Biol. Evol.*, 14(3):230–238, 1997.

- G. Barton. *Elements of Green Functions and Propagation: Potentials, Diffusion and Waves*. Clarendon Press, Oxford, 1989.
- George I. Bell and Jerzy Jurka. The length distribution of perfect dimer repetitive DNA is consistent with its evolution by an unbiased single-step mutation process. *J. Mol. Evol.*, 44:414–421, 1997.
- Bernd Brinkmann, Michael Klintschar, Franz Neuhuber, Julia Hühne, and Burkhard Rolf. Mutation rate in human microsatellites: Influence of the structure and length of the tandem repeat. *Am. J. Hum. Genet.*, 62:1408–1415, 1998.
- Graham J. R. Brock, Niall H. Anderson, and Darren G. Monckton. *Cis*-acting modifiers of expanded CAG/CTG triplet repeat expandability: associations with flanking GC content and proximity to CpG islands. *Hum. Mol. Genet.*, 8(6):1061–1067, 1999.
- Han G. Brunner, Gert Jansen, Willy Nillesen, Marcel R. Nelen, Christine de Die, Chris J. Howeler, Bernard A. van Oost, Be Wieringa, Hans-Hilger Ropers, and Hubert Smeets. Reverse mutation in myotonic dystrophy. *New Engl. J. Med.*, 328(7):476–480, 1993.
- Peter Calabrese and Rick Durrett. Dinucleotide repeats in the drosophila and human genomes have complex, length-dependent mutation processes. *Mol. Biol. Evol.*, 20(5):715–725, 2003.
- Peter P. Calabrese, Richard T. Durrett, and Charles F. Aquadro. Dynamics of microsatellite divergence under stepwise mutation and proportional slippage/point mutation models. *Genetics*, 159:839–852, 2001.
- JD Cleary, K Nichol, YH Wang, and Pearson CE. Evidence of *cis*-acting factors in replication-mediated trinucleotide repeat instability in primate cells. *Nat. Genet.*, 31(1):37–46, 2002.
- Pierre Gilles de Gennes. *Scaling Concepts in Polymer Physics*. Cornell University Press, Ithaca, NY, 1979.
- A. Di Rienzo, A. C. Peterson, J. C. Garza, A. M. Valdes, M. Slatkin, and N. B. Freimer. Mutational processes of simple-sequence repeat loci in human populations. *Proc. Natl. Acad. Sci. USA*, 91(3166–3170), 1994.
- Hans Ellegren. Heterogeneous mutation processes in human microsatellite DNA sequences. *Nat. Genet.*, 24:400–402, 2000a.
- Hans Ellegren. Microsatellite mutations in the germline: Implications for evolutionary inference. *Trends Genet.*, 16(12):551–558, 2000b.
- Nancy N. FitzSimmons, Craig Moritz, and Stephen S. Moor. Conservation and dynamics of microsatellite loci over 300 million years of marine turtle evolution. *Mol. Biol. Evol.*, 12(3):432–440, 1995.
- John Carlos Garza, Montgomery Slatkin, and Nelson B. Freimer. Microsatellite allele frequencies in humans and chimpanzees with implications for constraints on allele size. *Mol. Biol. Evol.*, 12(4):594–603, 1995.
- David B. Goldstein, Andres Ruiz Lares, Luigi Luca Cavalli-Sforza, and Marcus W. Feldman. An evaluation of genetic distances for use with microsatellite loci. *Genetics*, 139:463–471, 1995.
- Mário Gomes-Pereira, M. Teresa Fortune, Laura Ingram, John P. McAbney,

- and Darren G. Monckton. *Pms2* is a genetic enhancer of trinucleotide CAG-CTG repeat somatic mosaicism: Implications for the mechanism of triplet repeat expansion. *Hum. Mol. Genet.*, 13(16):1815–1825, 2004.
- Peter Harper. *Myotonic Dystrophy*. W. B. Saunders, Philadelphia, third edition, 2001.
- Peter Harper and Lesley Jones. Huntington’s disease: Genetic and molecular studies. In Gillian Bates, Peter Harper, and Lesley Jones, editors, *Huntington’s Disease*, volume 45 of *Oxford Monographs on Medical Genetics*, pages 113–158. Oxford University Press, Oxford, third edition, 2002.
- Bettina Harr and Christian Schlötterer. Long microsatellite alleles in *Drosophila melanogaster* have a downward mutation bias and short persistence times, which cause their genome-wide underrepresentation. *Genetics*, 155:1212–1220, 2000.
- Qing-Yang Huang, Fu-Hua Xu, Hui Shen, Hong-Yi Deng, Yong-Jun Liu, Yao-Zhong Liu, Jin-Long Li, Robert R. Recker, and Hong-Wen Deng. Mutation patterns at dinucleotide microsatellite loci in humans. *Am. J. Hum. Genet.*, 70:625–634, 2002.
- Manfred Kayser, Lutz Roewer, Minttu Hedman, Lotte Henke, Jürgen Henke, Silke Brauer, Carmen Krüger, Michael Krawczak, Marion Nagy, Tadeusz Dobosz, Reinhard Szibor, Peter de Knijff, Mark Stoneking, and Antti Sajantila. Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs. *Am. J. Hum. Genet.*, 66:1580–1588, 2000.
- Semyon Kruglyak, Richard T. Durrett, Malcolm D. Schug, and Charles F. Aquadro. Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. *Proc. Natl. Acad. Sci. USA*, 95:10774–10778, 1998.
- Semyon Kruglyak, Richard Durrett and Malcolm D. Schug, and Charles F. Aquadro. Distribution and abundance of microsatellites in the yeast genome can be explained by a balance between slippage events and point mutations. *Mol. Biol. Evol.*, 17(8):1210–1219, 2000.
- Yinglei Lai and Fengzhu Sun. The relationship between microsatellite slippage mutation rate and the number of repeat units. *Mol. Biol. Evol.*, 20(12):2123–2131, 2003a.
- Yinglei Lai and Fengzhu Sun. Microsatellite mutations during the polymerase chain reaction: Mean field approximations and their applications. *J. Theor. Biol.*, 224:127–137, 2003b.
- Yinglei Lai, Deepali Shinde, Norman Arnheim, and Fengzhu Sun. The mutation process of microsatellites during the polymerase chain reaction. *J. Comp. Biol.*, 10(2):143–155, 2003.
- Randell T. Libby, Darren G. Monckton, Ying-Hui Fu, Refugio A. Martinez, John P. McAbney, R. Lau, David D. Einum, K. Nichol, Carol B. Ware, Louis J. Ptacek, Christopher E. Pearson, and Albert R. La Spada. Genomic context drives SCA7 CAG repeat instability, while expressed SCA7 cDNAs are intergenerationally and somatically stable in transgenic mice. *Hum.*



- Mol. Genet.*, 12(1):41–50, 2003.
- Kateryna D. Makova, Anton Nekrutenko, and Robert J. Baker. Evolution of microsatellite alleles in four species of mice (genus *Apodemus*). *J. Mol. Evol.*, 51:166–172, 2000.
- Sergey M. Mirkin and Ekaterina V. Smirnova. Positioned to expand. *Nat. Genet.*, 31(1):5–6, 2002.
- Darren G. Monckton, Lee-Jun C. Wong, Tetsuo Ashizawa, and C. Thomas Caskey. Somatic mosaicism, germline expansions, germline reversions and intergenerational reductions in myotonic dystrophy males: Small pool PCR analyses. *Hum. Mol. Genet.*, 4(1):1–8, 1995.
- Rasmus Nielsen. A likelihood approach to populations samples of microsatellite alleless. *Genetics*, 146:711–716, 1997.
- Paul C. Painter and Michael M. Coleman. *Fundamentals of Polymer Science*. Technomic, Lancaster, Pa., second edition, 1997.
- CE Pearson. Slipping while sleeping? Trinucleotide repeat expansions in germ cells. *Trends Mol. Med.*, 9(11):490–495, 2003.
- Christopher E. Pearson, Kerrie Nichol Edamura, and John D. Cleary. Repeat instability: Mechanisms of dynamic mutations. *Nat. Rev. Genet.*, 6:729–742, 2005.
- Elisabeth Piñeiro, Laura Fernández-López, Josep Gamez, Ricard Marcos, Jordi Surrallés, and Antonia Velázquez. Mutagenic stress modulates the dynamics of CTG repeat instability associated with myotonic dystrophy type 1. *Nucl. Acids Res.*, 31(23):6733–6740, 2003.
- C. R. Primmer, N. Saino, A. P. Møller, and H. Ellegren. Unraveling the processes of microsatellite evolution through analysis of germ line mutations in barn swallows *Hirundo rustica*. *Mol. Biol. Evol.*, 15(8):1047–1054, 1998.
- Craig R. Primmer and Hans Ellegren. Patterns of molecular evolution in avian microsatellites. *Mol. Biol. Evol.*, 15(8):997–1008, 1998.
- C. Schlötterer, R. Ritter, B. Harr, and G. Brem. High mutation rate of a long microsatellite allele in *Drosophila melanogaster* provides evidence for allele-specific mutation rates. *Mol. Biol. Evol.*, 15(10):1269–1274, 1998.
- Deepali Shinde, Yinglei Lai, Fengzhu Sun, and Norman Arnheim. *Taq* DNA polymerase slippage mutation rates measured by PCR and quasi-likelihood analysis:  $(CA/GT)_n$  and  $(A/T)_n$  microsatellites. *Nucl. Acids. Res.*, 31(3):974–980, 2003. DOI: 10.1093/nar/gkg178.
- Mark D. Shriver, Li Jin, Ranajit Chakraborty, and Eric Boerwinkle. VNTR allele frequency distributions under the stepwise mutation model: A computer simulation approach. *Genetics*, 134:983–993, 1993.
- Richard M. Sibly, John C. Whittaker, and Melanie Talbot. A maximum-likelihood approach to fitting equilibrium models of microsatellite evolution. *Mol. Biol. Evol.*, 18(3):413–417, 2001.
- Richard M. Sibly, Andrew Meade, Nicola Boxall, Michael J. Wilkinson, Dave W. Corne, and John C. Whittaker. The structure of interrupted human AC microsatellites. *Mol. Biol. Evol.*, 20(3):453–459, 2003.
- Hidenori Tachida and Masaru Iizuka. Persistence of repeated sequences that

- evolve by replication slippage. *Genetics*, 131:471–478, 1992.
- John C. Whittaker, Roger M. Harbord, Nicola Boxall, Ian Mackay, Gary Dawson, and Richard M. Sibly. Likelihood-based estimation of microsatellite mutation rates. *Genetics*, 164:781–787, 2003.
- Xin Xu, Mei Peng, Zhian Fang, and Xiping Xu. The direction of microsatellite mutations is dependent upon allele length. *Nat. Genet.*, 24:396–399, 2000.
- Z Yang, R Lau, JL Marcadier, D Chitayat, and CE Pearson. Replication inhibitors modulate instability of an expanded trinucleotide repeat at the myotonic dystrophy type 1 disease locus in human cells. *Am. J. Hum. Genet.*, 73(5):1092–1105, 2003.